DOCUMENT RESUME

ED 047 913 RE 003 380

AUTHOR Tuinman, J. Jaap; Blanton, B. Elgit
TITLE The Cloze Procedure: Corollaries of a Search Hypothesis.
INSTITUTION Indiana Univ., Bloomington. School of Education.
PUB DATE Feb 71
NOTE 13p.

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS *Cloze Procedure, *Correlation, *Hypothesis Testing, Junior High School Students, Language Research, Reading Comprehension, Reading Materials, *Reading Processes, *Reading Research, Reading Tests

ABSTRACT

In order to determine whether the kind of process underlying cloze responses is indeed a systematic and exhaustive search, a study was conducted exploring some corollaries to such a search hypothesis. It was assumed that subjects would generate responses representing a number of word types, that some of these word types would be sensible and some nonsensical, and that responses would be representative of the entire body of possible response words. Five versions of a 300-word cloze passage, every fifth word deleted, were administered to 390 junior-high-school students who were randomly assigned to one of the versions. Protocols were hand scored and success probabilities were calculated. A correlation matrix among seven variables was calculated and analyzed using a stepwise regression program. Significant correlations were noted among the seven variables, with the highest correlation appearing between size of response body related to success probability. It was concluded that the general search hypothesis appeared to be sustained since distribution of responses was related to success probability and since the ratio of nonsense to sensible responses was relevant to that distribution. It was implied that a search process could be characterized as systematic in part. Tables and references are included. (MS)

ED047913

The Cloze Procedure

Corollaries of a Search Hypothesis

J. Jaap Tuinman
Reading Program
Institute for Child Study

B. Elgit Blanton
Reading Program
ERIC/CRIER

Indiana University

Students of the processes underlying the acquisition of information from written communication have from time to time used a technique referred to as the cloze procedure. So far, the cloze procedure has been used predominantly in connection with the measurement of reading achievement and readability formulas. Rankin (1965) and more recently Bickley, Bickley, and Ellington (1970) have summarized research dealing with the cloze procedure and its applications.

A cloze task consists of a language passage in which words have been deleted according to some prearranged scheme. Subjects are asked to guess the missing words. Although various scoring techniques have been described (e.g., Taylor and Waldman, 1969), in most cases a right-wrong scoring procedure is used where exact replacements of deleted words constitute correct responses.

The question asked then is: what kind of process underlies the production of cloze responses? Most popular has been the assumption that the organism engages in some kind of systematic and exhaustive search process. One bit of evidence for such a search hypothesis was provided by Taylor (1954). He found that the number of word types emitted by a sample of subjects in response to a particular deletion correlated highly negatively with the probability of that deletion being "clozed" successfully.

U. S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS

RE003 380

Taylor, who conceived the cloze procedure (Taylor, 1953), together with some colleagues studied the relative latencies of semantic aphasics, stutterers, and normals for cloze items requiring unique or non-unique responses. (Taylor, Lore and Waldman, 1967). Unique responses were responses to blanks which were constrained by the bilateral context to the point that only one specific word could possibly make sense. One, not unexpected finding of this study was that unique responses required shorter latencies than non-unique responses. This result would be predicted if one were to assume that a systematic search process underlies the production of closures.

In Weaver's (1965, p. 131) opinion, the constraints involved in the cloze "enable us to get a close-up view of what is occurring at particular points in language passage." A major issue in this context is the nature of what exactly is occurring when, in the midst of a decoding operation, a reader is forced to engage in a productive operation.

The solution of this issue is important both for a theory of reading and of language processes in general. In normal reading, or listening for that matter, very little interruption of the decoding process from the outside takes place. Conceivably, however, there are many instances in both these receptive processes where internally stimulated productive behavior interrupts the decoding process per se. The degree to which this is true seems to determine the importance of understanding the nature of the cloze task for an increased understanding of the nature of reading process.

Weaver (1965, p. 130) challenges the postulation of this kind of search hypothesis to some extent: "It is easy to show that exhaustive search procedures would be much more time consuming than any human being could

afford, and the illogicalness of many of our search efforts is obvious." Thus, Weaver considers the positioning of a logically exhaustive search process as only an approximation to the situation in reality.

To date, hardly any data have been presented which throw a direct light upon the degree to which extent one can speak of search behavior in connection with cloze. In the present study, some of the corollaries to a search hypothesis were explored in a tentative fashion. The following assumptions were made:

a) Given a cloze blank, a sample of n subjects will generate n responses (word tokens) which represent k word types where $o \leq k \leq n$. The minimal value of k occurs when no subject attempts a response; k is maximized when all Ss emit a different response.

b) The k response types consist of two different kinds. When considering the fill-ins to a particular cloze-blank one is always struck by the fact that some responses just don't make sense at all; they are either syntactically inadmissible words or seem semantically incongruous with the context. These are the words which fall in the N(onsense) class; the others are S(ensible) responses. Theoretically at least one can assume that the k response types consist of k_1 N-types and k_2 S-types, where $k=k_1+k_2$.

c) The sample of responses emitted by a given sample of subjects is representative of a population of responses for a particular blank for the population of subjects from which that sample was drawn. That is, the researcher never has data about the corpus of words from which his Ss supposedly sample unless he assumes that the words actually emitted are representative of that corpus, or if you will, population of words.

In regard to assumption b, one further comment needs to be made. Consider the following sentence:

The man _____ his house.

Consider now the following set of responses emitted by n = 10 Ss.

bought (2), painted (3), sold (1), liked (2), was (1), embraced (1).

If one assumes that these word types do indeed represent the corpus of words through which subjects search when attempting to fill in the gap, the following observations seem in order. (1) The first four words seem to belong in the S-class while the other two words seem to be N words. (2) Note that it is very difficult to make up one's mind about which of the S-words is probably the correct response. The decision in regard to the two other words is much easier.

While it is most likely a simplification, it seems temporarily defensible to assume that, if a systematic search process takes place, the total number of decisions regarding the rejection or accepting of a word as correct choice equals $k_1 + a$ where a is a value based on the possible permutations among k_2 S-class elements. This of course begs the question in regard to equal attractiveness of all S-words. However, presumably $a > k_2$. Furthermore, it seems reasonable to assume that given a fixed number of total responses ($=n$) an increase in reject-accept decisions means a decrease in the probability of guessing the right word.

It was mentioned above that the size of the distribution of responses to a particular deletion was shown to be related to the probability of a correct answer. The speculations above, interpreted as a corollary to a systematic search hypothesis seem to indicate that the distribution of N and S words within the k response types also might affect the success probability. That is: it is hypothesized that both these parameters are determinants of success probability. To state this hypothesis differently: a regression equation of the form $p = a - b_1 k + b_2(N/S)$ was postulated, where

p = proportion of Ss correctly filling in the blank

k = sample size

N/S = the ratio of N-type responses and S-type responses.

Finally, it is recognized that whereas presence of the relationship hypothesized conceivably admits both of a logical and exhaustive search and of a more heuristic procedure absence of such a relation seems more damaging for the former.

Procedure

Materials. The data analyzed here were collected by administering five versions of a 300 word cloze passage in which every fifth word had been deleted. The passage was taken from a junior high school reading text. The versions differed in the first and therefore in the subsequent words deleted. Across the five versions all 300 words appeared as blanks once. The passage was preceded and followed by paragraphs of respectively 140 and 100 words long.

Subjects. Ss were 390 junior high students, nearly equally divided over the 7th, 8th, and 9th grades.

Procedure. The Ss were randomly assigned to one of the five cloze versions. The task was explained to them by means of an illustrative paragraph. They were then asked to "read the story and fill in the exact words which you think were left out."

Analysis. The protocols were hand scored and for each word the success probability (number of Ss guessing the word) was calculated. In the analyses presented below only the nouns (n = 51) are included. For these 51 nouns the following statistics were calculated from the response distribution of each: (1) k (= total word types); (2) k_1 (= word types

in the N-class); (3) k_2 (= word types in the S-class); (4) $p_1 = k_1/k_2$; (5) $p_2 = rk_1/rk_2$ where rk_1 stands for the total number of responses in the N-class and rk_2 for the total number of responses in the S-class; (6) p (= proportion of correct responses). It must be noted that the classification of word types as either N or S is subjective. In the majority of the cases, however, classificatory judgments were rather unambiguous. The variables included in the analysis are summarized in Table 1.

Insert Table 1 about here

In preliminary analyses additional variables were included but dropped because of redundancy. They were: k_1/k; k_2/k; rk_1/m; rk_2/m; k_1/rk_1; k_2/rk_2.

A matrix of correlations among the variables was calculated and analyzed using the RMD 02R, Stepwise Regression program.

Results

Table 2 presents the correlations among the 7 variables included in analysis. The Spearman-Brown reliabilities of the five Cloze versions ranged from .82 - .91 with a .89 median.

Insert Table 2 about here

A few remarks are in order. First of all, it seems clear that k, the size of the corpus in terms of word types is highest related to the success probability. The more word types emitted the smaller the probability of success for a specific closure. This result simply confirms Taylor's findings in this respect mentioned above. Taylor found a -.87 rank order correlation between p and an information statistic calculated on the basis of the number of word types emitted and the frequency with which each word

was chosen by the total sample of Ss. It now appears that a large portion of this correlation can be explained by corpus size <u>per se</u>.

In addition it may be noted that k_1 and k_2 differ vastly in their relationship to p_2 (= rk_1/rk_2). The reason for this is that whereas k_1 is highly related to rk_1 (r = .89, not shown in Table 2), k_2 is not related to rk_2 (r = .24, not shown). This simply means that in the case of N-words the number of word types varies closely with the number of word tokens: not many N-word types attracted more than one respondent.

In order to further explore the relationship of word corpus characteristics to p, the probability of successfully clozing the deletion, two regression equations were computed. First, all variables were included in the calculations. The resulting equation was:

$$p = .82198 - .02412k + .02458p_1,$$

where p = proportion of subjects correctly filling in a blank

k = the size of the distribution of word types at the point of that blank and

p_1 = the ratio of N and S word types.

It may be noted that the direction of the regression coefficients is in the anticipated directions. A multiple R of .747 is associated with this equation. The standard error of estimate is .85. The inclusion of the first variable accounts for 50 per cent of the variance; the second variable adds 6 per cent. The F-ratio associated with the proportion accounted for by regression equals 30.33 (df = 4,48) After inclusion of these two variables, no other variables possessed significant partial correlations with the criterion. Parenthetically, it may be noted that the simple correlation between p_1 and p was only .04. However, the

correlation between these variables with the effect of corpus size (k) partialled out increased to .34.

In terms of the theoretical issue underlying this study an interesting question remains unanswered by the above results. Apparently success probability (p) is highly related to corpus size. However, given a corpus of a specific size, what characteristics of the corpus do determine p? A complete answer to this question would indeed shed a great amount of light on what processes occur when a subject is faced with a specific blank for which (as is theoretically always the case) the corpus from which he selects is fixed in size.

Presumably, this question can best be researched by studying the variables mentioned above, quite possibly in conjunction with other variables, for a number of blanks with identical corpus sizes. To generate such data is rather difficult and costly. Out of curiosity the authors reanalyzed their data removing the size variable k from the system of correlations. Again using stepwise regression the following equation resulted:

$$p = .77250 - .0143k_1 + .06439p_2 - 1.56618p_4$$

where p = proportion of subjects correctly filling in a blank

k_1 = size of the distribution of N words

p_2 = the ratio of the N and S word tokens

p_4 = the ratio of S word types and word tokens

The resulting multiple R equals .719 (Std. error of est. - .196). The variables were entered in the equation in the following order: p_4, k_1, p_2 accounting for respectively 42, 5, and 5 per cent of the variance. ($F = 16.79$ with 3 and 47 df.) The total percentage accounted for (52%) compares not too unfavorably with that of the first equation (56%).

Discussion

The production of cloze responses is a highly complex activity, most likely involving utilization of a great many syntactic and semantic clues alternately at the level of conscious consideration of alternatives and automatically acting.

One difficulty in interpreting the present findings in terms of their explanatory contribution is the concept of the corpus of words emitted as an approximation to an hypothetical corpus of words being searched for the correct response. Much uncertainty exists as to the conditions under which the overt corpus may be taken to be representative of its postulated covert counterpart. Are, for instance, too stringent time constraints when performing the task related to distortion of representativeness of the sampled corpus? Questions such as this need eventually to be answered in order to achieve a satisfactory description of the process of closure.

At this point it seems fair to say that the corollary derived from a general search hypothesis seems sustained. Not only is distribution related to p (as would be expected on the basis of any kind of search hypothesis), also relations pertinent to the ratio of N words (or decision reducing elements) and S words (or decision increasing elements) is relevant. This seems to be borne out by both regression solutions. As mentioned above, the equations obtained admit both of a systematic and a heuristic search procedure. Only nonsignificant regression coefficient of all variables related to the N/S ratio could be taken as evidence, however weak, of absence of systematic search. Presumably, it makes only sense to speak of decision reducing or decision increasing elements if some kind of comparison of word tokens prior to emitting a response takes place. In the light of the evidence, it seems unlikely that at no point in the response formation period such systematic

comparison is engaged in by the responding organism. This is a contra-indication of a completely heuristic procedure.

For the moment, it appears that experimentation is needed to reveal the extent or nature of organization in the search process. At this point it seems not unlikely that the search process can be characterized as systematic in part. That is, while a part of the time spent in searching for the correct solution may be used for heuristic searching, this does not necessarily exclude theoption at a particular moment in the search process to revert to a much more careful, deliberate and systematic analysis of the various choices available.

Table 1

Description of Variables

Variable	Description
k	number of total word types emitted
k_1	number of total word types of class N, emitted
k_2	number of total word types of class S, emitted
P_1	ratio of N and S word types
P_2	ratio of N and S word tokens
P_3	ratio of N word types and N word tokens
P_4	ratio of S word types and S word tokens
p	proportion of subjects correctly falling in a given blank

Table 2

Matrix of Correlations among All Variables*

	(1) k	(2) k_1	(3) k_2	(4) P_1	(5) P_2	(6) P_3	(7) P_4	(8) P
(1)	1.00	.71	.54	.28	.38	-.08	.78	-.71
(2)		1.00	-.19	.73	.61	-.38	.31	-.41
(3)			1.00	-.48	-.18	.36	.74	-.48
(4)				1.00	.45	-.27	-.14	-.04
(5)					1.00	-.36	.49	-.26
(6)						1.00	.00	.10
(7)							1.00	-.65

* for any $r \geqslant .25$, $p \leqslant .05$,
for any $r \geqslant .36$, $p \leqslant .01$

References

Bickley, A. C., Ellington, B. J. and Bickley, R. T. The Cloze Procedure: A Conspectus. Journal of Reading Behavior, 1970, 2, 232-249.

Rankin, E. F. Cloze procedure--a survey of research. In E. Thurston and L. E. Hafner (Eds.) Fourteenth Yearbook of the National Reading Conference, Milwaukee, 1965.

Taylor, W. L. Cloze Procedure: a new look for measuring readability. Journalism Quarterly, 1953, 30, 414-433.

Taylor, W. L. Application of "cloze" and entropy measures to the study of contextual constraint in samples of continuous prose. Unpublished doctoral dissertation, University of Illinois, 1954.

Taylor, W. L.; Lore, J. I., and Waldman, I. N. Relative latencies of semantic aphasics, stutterers, and normal controls for "cloze" items requiring "unique" and "non-unique" oral responses. Paper presented at the annual meeting of the Academy of Aphasics, Ann Arbor, Mich., 1967.

Taylor, W. L. and Waldman, I. N. Latency and Focus methods of cloze quantification. Paper presented at the annual meeting of the National Reading Conference, Atlanta, 1969.

Weaver, W. W. Theoretical aspects of the cloze procedure. In E. L. Thurston and L. E. Hafner, (Eds.) Fourteenth Yearbook of the National Reading Conference, Milwaukee, 1965.